



SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Treco, Douglas A.
Heartlein, Michael W.
Hauge, Brian M.
Selden, Richard F
- (ii) TITLE OF INVENTION: Protein Production and Delivery
- (iii) NUMBER OF SEQUENCES: 30
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.c.
 - (B) STREET: Two Militia Drive
 - (C) CITY: Lexington
 - (D) STATE: Massachusetts
 - (E) COUNTRY: USA
 - (F) ZIP: 02173
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08,406,030
 - (B) FILING DATE: 17-MAR-1995
 - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/243,391
 - (B) FILING DATE: 13-MAY-1994
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/985,586
 - (B) FILING DATE: 03-DEC-1992
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/911,533
 - (B) FILING DATE: 10-JUL-1992
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/787,840
 - (B) FILING DATE: 05-NOV-1991
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 07/789,188
 - (B) FILING DATE: 05-NOV-1991
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: PCT/US93/11704
 - (B) FILING DATE: 02-DEC-1993
- (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER: PCT/US92/09627
 - (B) FILING DATE: 05-NOV-1992

(viii) ATTORNEY/AGENT INFORMATION:
 (A) NAME: Granahan, Patricia
 (B) REGISTRATION NUMBER: 32,227
 (C) REFERENCE/DOCKET NUMBER: TKT95-01

(ix) TELECOMMUNICATION INFORMATION:
 (A) TELEPHONE: (617) 861-6240
 (B) TELEFAX: (617) 861-9540

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AATTGCTCCT CGTGGTCATG CTTCT

25

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 21 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CTGTGAAGGA CATGGGAGTC A

21

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 4488 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TCTAGAGTCA GGATGGCACT GAAGGTCTCT GGGGAAGGGA CGATGATGAG AGCCCCGTCAG

60

AAACCCTCCC	CCCTTTCCTG	GGTGATAGAG	AAGACTCAGA	ACTTCACGCC	CGGGGCTCTT	120
TGCTCCCTAC	CTGCAGCCAG	GGCCCCGTGC	GATGAGAGCC	CCCAGACCTC	CCTGAAGGGT	180
GAGTGAGTGT	CACAAGTGCC	ACATGCAGCT	GTTCTGCCCT	AAGGAGCCGC	AGAGACAACC	240
GAGGCACTGC	CCGCCACACC	CCACAGACCT	GGAGCAGAGA	GACAAGAAGG	CCCTACGCTC	300
AGACACTGTG	CAGGCTAGGC	CAATTAGGAT	GCCCAGGCAG	GGCTTATGAA	AAAGGAACAT	360
GGAAAGGAAC	CTCCAGGGTG	CCCTAGGAAG	CTTAAGAAAG	AACGCTGGAG	CCAGATGCTT	420
GGGTTCCAAT	CCTGGCTGCA	CCACTTCCTA	GCTGTGTGAC	CTTGAATCAA	ATCACATTAT	480
CCTACTGAGC	CTCAGTTCCT	CCTTCTGTAA	AATGGGCATC	ATAATGTCAG	TGCCTTCCTC	540
CCACTGGGCT	GTGGTGAGGA	CCACGGGAGG	CAATGCAGAG	CATGCTCTCG	GCACAGTGCC	600
CAGACTGGGC	AAGTGCTATA	AATGGCATCA	TCTCACCAGG	CCTATCTTGG	GTTGRGTGGG	660
CTGCAGGGTG	CTCAAACAGG	ACACTGCCAT	TGGAGTCTGA	GAAGCGGATC	CTGGTAGGGC	720
GGTCCAGCCT	GGGAATGAGA	GGTCGGGTGA	GGCCGGACTG	AGCCAAAAGC	AGCCCCCTCC	780
AGCTCTCCCA	GTTTCCCTCC	SGGCCCCGGC	AGCGTGACCC	CTCCTTGCTC	CTTCCCCTTT	840
CTCACCGCCT	GTAGGAGATA	GAGAAGCGGA	GGCTAGAGCG	CCAGCAGCGA	GACTCGGCTC	900
GTGCCACCGC	CTGCGACCTC	GGCCCTGTCA	GCAGCGCCAC	GAAGTCTGGG	ACGGGAGGAA	960
GATGGCCTGA	GCACTGTCAA	ACGCCGCTTT	GGTGGCCCAG	CCTCAACCAC	AACCCCGCTG	1020
TTCGCCAGCC	CCCTACCCGT	GTGGCCGTCA	CCACGGGCCC	GCTCCTCAGC	GCCTGGCTCC	1080
CCGCGGTGCG	TATAACTGCG	ATGCTCCGGG	TCCC CGGGAT	ACACGAAGGA	CAGGCCGCTC	1140
GGCTGCCGCT	CCGAAGTGCT	GCGCTCTGCG	GSGGGGGGGT	AAGAACACGG	GCTTCAGCTG	1200
GCCATGGGAA	AGGCCAGTCC	GACGCCCCAT	CCAAGTGGCC	CGGGACCTAG	TATCGTGGCC	1260
CTGCCTCCCT	CCCCGCAGCG	GAGCAAGACT	TACCCTGGGG	GCAGGTCTGG	CAGCAGTGTC	1320
COGGCAGCTG	GCGCGGCTGC	CCACAGGCCG	GGGTGGGCA	CTCTGGTTTG	ATGTTCTTGC	1380
AGCTGACCCT	GCCAGGCCCC	TGGTACGGCG	ACCCCACTGA	GGCTGCTCCC	GGAAAAGGCG	1440
GGAAACCCAA	GTGAGTGCAA	GATGCCAACT	GATGAGACCC	CCCCAGGCAA	GGATGTCCCG	1500
CAGAGTCAGC	CAGCTCTGCC	ACTTACAAGC	TGCGTGACCC	TAGACAAGCT	ACTTCATCTC	1560
TCTGGGCCTC	AAGGTCCCTG	TCTGGAAAAT	GGGGATAATA	ATACTCTCTA	TCTAGCAAGG	1620
CTGCCATGAG	AGTTAGATGA	GCAGGGAACG	AAACGGAGTT	GGCACAGAGC	CTCACACAGA	1680
GTGGGCGATC	AGTAACAGCA	CCTAAGAATT	GGAGGGGCTG	ATTCCCCTTC	CTCCACCAGA	1740
AAAATATCCC	CAACATCTGC	CGACTGGGCT	CCTTCTCAGC	AGCTCCGAGT	CCACTCCGAC	1800
GGCCGCGCGA	CCCGGCCGTC	CCCACCCGCC	AGCCCGGGCC	GGCCGCGGGG	TGCACTCACC	1860
GCCTCGCAGG	CCACAGCACG	CAGCGCATCA	CCCGAATGG	CTCCCCTAGG	TCCGGGTGCC	1920

ACGTCTCGTC	CAAGGCATAG	ACCTTCCCCG	CGAAGTGCAG	CCTGCGGGAC	GGGCTTGGCT	1980
GGAGGCGCTG	CCCAGCTCGC	GCCGTGTGCC	GCCCCGGGGG	CTGCCCCGGG	GTCCCGGGTC	2040
CCAGGCACCG	CGCCCTTCTG	CCCCCGCCCA	CCCTCCGGGC	CGCCCGCCGC	GCCGAGCCAC	2100
CTGCGCCCCG	CGCCCCTCCT	CCGGCTCGGC	TGACTCGCCC	CGAGCCCGAC	TCCCCGCCCC	2160
CCTCCCCCGG	GCGCCACCT	ACCCTGCTGC	CCGAACGGGC	AGCGGCTCCT	TCTCAGAACG	2220
GATGGGCAGC	ACGGGGGCTC	TCGGGCCGGC	CGGGGCGGGA	GCCGAGCAGC	AGCAGCCCCA	2280
GGAGCAGCAG	CGGGGCCGGC	GGGGCCGGGA	GGGCHCGGCA	TGACGCGAAC	GGGACAGCTG	2340
GGGAGGAGGG	AGGGAGGAGG	GCGCGGGAGC	GGGCGGAGGG	AGGGAGGCGG	GAGTGC GGAG	2400
GGCGGAGGGC	CGGGCCGGGG	GCGGTGCGGC	GGGAGGGGGC	CGGGGCCGGG	GCCGGGGCCG	2460
GGGCAGTGCC	CGCGAGGGGC	TCGTGCGGCG	GCCGCAGAGT	CGGCGCCGGG	CCGGGGCGGG	2520
CGGGAGGAGC	GGCCGGGAGG	AGCGCGGGCG	GGCGGGCGCT	GACCCGGGCC	GTACGCGGCT	2580
CTACTGCCCC	GGGCGCCGGC	TCCGGCCCCG	TTTATGCCCC	GCGCCCGACG	CCCCGGCCGG	2640
GGGCCTCCTC	CTCAGCAAAC	GGGGCGGCGG	CGGCGGCTCG	GCGAGGGGCC	GCTGAGCCCC	2700
GGGGGTCCGA	CCCAGCAGCA	GCGGCCCGGA	TCGCGGGTGG	GGGAGGGGAG	GGAGGGCTGG	2760
GACCCGGGCAG	GGGAGGAGGG	AGGGGCGGGA	GGGGAAGGGG	GAGCGGGGGA	GGGGGAGGGG	2820
AGGGACCAGG	GGGCGCGAAG	AGGGGGAGGA	GAGGCGGCCC	GGAGCCCCCG	CTGCTGGCGG	2880
CCACAGGGCG	GCTGGACCAG	GAGGTGCGTG	TCCAGCCCAG	GAAGGGAGCC	TCAGGCTAGG	2940
GAGGGGCAGA	GGCTTACCTG	AGGCCTGGAC	CGCTCTGTGA	GCGAGGCCCC	GTTCCGCCCC	3000
AAGGATAAAC	TTGTCTTTAA	AGATACACGT	ACAGGAAAGG	TCCATCAGCC	GATCTCCCCC	3060
TGCCTGGGCC	CACAGCGCCC	CCCAAACCCT	CACCACCCTC	TCTCACTGCC	TAGCCTGCCT	3120
CCCTACCTTC	TCTCTGAGGT	CGCTCCTCWT	TCTTGTGTGA	CCCAGRACAG	GGACCTAGCC	3180
AGAAACCGGC	AGCATTCCCC	CTTCTGTGGA	GTGACAGTAT	CTCCCTCTCA	TTGTAACTTA	3240
TCCTCAGGCG	CATTGACAG	TCCCCTCTTG	CTTTCTCACC	CCCTTCCTTC	ACCCAAGGGA	3300
CCCTCTGCCT	CTCCAGCCCA	CTCCCAGCCT	CCTTTCTCTT	GGTCCCTGG	TCATGCCTGC	3360
CTCCCTGTCT	CCTGTCTCTC	CCTCCCACAC	ACACCCACTA	TCCTCCCAGC	TATCCCAGCA	3420
CCCTCCTTCC	TAATCTTGGG	AGACATCTCG	TCTGGCTGGA	CGGGAAAATT	CCAGGATCTA	3480
GGCCACACTT	CTCAGCAGAC	ATGCCCATCC	TTGGGGAGGA	GGAACAGGAG	AGAGCCTGAG	3540
GAAGTTCTGG	GGGACAGGGG	GATGATGGGA	TCAAGGTCAG	GCCAGGAAGC	CCCTGAGGAC	3600
AGAGACTGTG	GGGAGACTTG	GGAAGGAAAG	GAGCTAGAGC	CAGGGCCAAA		3660
GGAAAAGGGG	GGCCAGCAGG	GWGGTATTTG	CGGGGGAGGT	CCAGCAGCTG	TCTTTCCTAA	3720
GACAGGGACA	CATGGGCCTG	GTTATTCCTC	TTGTCACATG	TGGAACGGTA	GGAGATGGAA	3780

GACGGAGACA GAACAAGCAA AGGAGGGCCC TGGGCACAGA GGTCTGTGTG TG TAGCCATC	3840
TAAGCCACTG GACCCAGCA GACGAGCACC TAAGCTCAGG CTTAACCAGT GCACGTGTGC	3900
GCACATACTG TGCCCCGCAC CTGACGTCCA CTCAACCCGT CCAAACCCCT TCCCCATAAC	3960
ACCAACCCAT AACAGGAGAT TTCTCTCATG TGGGCAATAT CCGTGTTCCT ACTTCGAAAG	4020
GGGGAATGAC AAGATAGGAC TCCCTAGGGG ATTACAGAAA GAAAAGCAGG AAAGCAAGCA	4080
TCCTGTTGGA TTTCAGCAGC AGGTATGATG TCCAGGGAAA AGAAATTTGG ATAGCCAGGG	4140
AGTGAAAACC CCACCAATCT TAAACAAGAC CTCTGTGCTT CTCCCCAGC AACACAAATG	4200
TCCTGCCAGA TTCCTCCTGG AAAAAACTTC TGCTCCTGTC CCCCTCCAGG TCCAGGTTCG	4260
CCATGTCCAG GAAAAGATGG ATCCCCCTCA TCCAAATCTT CTCCGTGTGT GCTGTGGGTG	4320
GAGTGAGTRG WARCCCTGGT CCAGGCAGGG VGCTCCAGGG AAGAGCAAGG CGTCACTTCC	4380
GGGSGCCTTC ACCAGTGTCT GGTGGCTCCC TTCTCTGATT GGGCAGAAGT GGCCAGGCA	4440
GAGCGTATGA CCTGCTGCTG TGGAGGGGCT GTGCCCCACC GCCACATG	4488

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2455 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TCTTCTACC CATCTGCTCC CCAGAGGGCT GCCTGCTGTG CACTTGGGTC CTGGAGCCCT	60
TCTCCACCCG GTGAGTGGCC AGCAGGGTGT GGGGTTATGT GAGGGTAGAA AGGACAGCAA	120
AGAGAAATGG GCTCCAGCT GGGGGAGGGG CAGGCAAACT GGAACCTACA GGCAGTACC	180
TTTGTGAGA AGAGTGTAGC CTTCCAGAA TGGGAGGAGC AGGGCAGAGC AGGGGTAGGG	240
GGTGGGGTGC TGKTTTCCTG AGGGACTGAT CACTTACTTG GTGGAATACA GCACAGCCCT	300
GGCTGGCCCT AAGGAAAGGG GACATGAGCC CAGGGAGAAA ATAAGAGAGG GAGCTGCACT	360
TAGGGCTTAG CAAACACAGT AGTAAGATGG ACACAGCCCC AATCCCCATT CTTAGCTGGT	420
CATTCCTCGT TAGCTTAAGG TTCTGAATCT GGTGCTGGGG AAGCTGGGCC AGGCAAGCCA	480
GGGCGCAAGG AGAGGGTAAT GGGAGGAGGG CCCACTCATG TTGACAGACC TACAGGAAAT	540
CCCAATATTG AATCAGGTGC AAGCCTCTTT GCACAACTTG TGAAAGGAGG AGGAAGCCAT	600
GTGGGGGGTC CTGTGAAGGA ACCGGAAGGG GTTCTGCCAA GGGGGCAGGG AGGCAGGTGT	660
GAGCTATGAG ACAGATATGT TAGTGGGCGC CTAAGACAAG GTAAGCCCCT AAGGTGGGCA	720

TCACCCAGCA GGTGCCCCGTT CCTGGGCAGC TGGTTTCAGG AAGGAAGTCC CAGAACTGTT	780
AGCCCATCTC TTGGCCTCAG ATAATGGAGT ATTTTCAGGAC TTGGAGTCCA GAGAAAAGCT	840
CCAGTGGCTT TATGTGTGGG GGTAGATAGG GAAAGAATAG AGGTTAATTT CTCCCATAACC	900
GCCTTTTAAT CCTGACCTCT AGTGGTCCCA GTTACAGCTT TGTGCAGTTC CCCTCCCCAG	960
CCCCACTCCC CACCGCAGAA GTTACCCCTC AACATATTGC GCCCCTTTC CAGTTCCTCA	1020
CCCAGGCCCT GCATCCCATT TTCCACTCTC TTCTCCAGGC TGAAGCCACA ATACTTTCCT	1080
TCTCTATCCC CATCCCAGAT TTTCTCTGAC CTAACAACCA AGGTTGCTCA GAATTTAAGG	1140
CTAATTAAGA TATGTGTGTA TACATATCAT GTCCTGCTGC TCTCAGCAGG GGTAGGTGGC	1200
ACCAAATCCA TGTCCGATTC ACTGAGGAGT CCTGACAAAA AGGAGACACC ATATGCTTTC	1260
TTGCTTTCTT TCTTTCTTTC TTTCTTTCTT TTTTTTTTTT GAGACGGAGT TTCACTCTTA	1320
TTGCCCAGGC TGGAGTGCAA TGGTGCGATC TCGGCTCACC ACAACCTCCG CCTCCCAGGT	1380
ACAAGCGATT CTCCTGTCTC AGCCTCCCAA GTAGCTTGGA TTACAGGCAT GAACCACCAC	1440
ACCCTGCTAG TTTTTTTGTA TTTCGTAGAG CCGGGGTTTC ACCATGTTAG TGAGGCTGGT	1500
GGCGAACTCC TGACCTCAGG TGATCCACCC GCCTTGGA CTCCAAAGTGC TGGGATTACA	1560
GGCATGAGCC ACTGCACCCG GCACACCATA TGCTTTTCATC ACAAGRAAAT GTGAGAGAAT	1620
TCAGGGCTTT GGCAGTTCCA GGCTGGTCAG CATCTCAAGC CCTCCCCAGC ATCTGTTTAC	1680
CCTGCCAGGC AGTCTCTTCC TAGAACTTG GTTAAATGTT CACTCTTCTT GCTACTTTCA	1740
GGATAGATTC TTCACCCTTG GTCCGCCTTT GCCCCACCCT ACTCTGCCCA GAAGTGCAAG	1800
AGCCTAAGCC GCCTCCATGG CCCCAGGAAG GATTGAGGGG AGAGGCCCCA AACAGGGAGC	1860
CACGCCAGCC AGACACCCCG GCCAGAATGG AGCTGACTGG TGAGAACACA CCTGAGGGGC	1920
TAGGGCCATA TGGAAACATG ACAGAAGGGG AGAGAGAAAG GAGACACGCT GCAGGGGGCA	1980
GGAAGCTGGG GGAACCCATT CTCCCAAAAA TAAGGGGTCT GAGGGGTGGA TTCCCTGGGT	2040
TTGAGGTCTG GGTCTGAAT GGAATTCCT GGAATACCAG CTGACAATGA TTTCTCCTC	2100
ATCTTTCAAC CTCACCTCTC CTCATCTAAG AATTGCTCCT CGTGGTCATG CTTCTCCTAA	2160
CTGCAAGGCT AACGCTGTCC AGCCCGGCTC CTCCTGCTTG TGACCTCCGA GTCCTCAGTA	2220
AACTGCTTCG TGAATCCAT GTCCTTCACA GCAGACTGGT GAGAACTCCC AACATTATCC	2280
CCTTTATCCG CGTAACTGGT AAGACACCCA TACTCCAGG AAGACACCAT CACTTCCTCT	2340
AACTCCTTGA CCCAATGACT ATTCTTCCCA TATTGTCCCC ACCTACTGAT CACACTCTCT	2400
GACAAGGATT ATTCTTCACA ATACAGCCCG CATTAAAAAG CTCTCGTCTA GAACT	2455

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TTTTGCGGCC GCTCGAGGAC ATTGATTATT GACTAGT

37

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 36 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

TTTTGCGGCC GCCGGTACTT ACGTCACTCT TGGCAC

36

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TTTTCTCGAG GACATTGATT ATTGACTAGT

30

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 60 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

CATGGGTCTT TTCTGCAGTC ACCGTCCTTG CTACCCATCT GCTCCCCAGA GGGCTGCCTG

60

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 60 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CAGGCAGCCC TCTGGGGAGC AGATGGGTAG CAAGGACGGT GACTGCAGAA AAGACCCATG

60

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TTTTGGGCCC TCCTCCCATC ACCCTCT

27

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TTTTCTCGAG GACATTGATT ATTGACTAGT

30

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CGCGGATTCC CCGTGCCAAG CCTAGCGGCA ATGGCTACAG GTGAGAACAC ACCTGAGGGG	60
CTAGGGGCCA	69

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 69 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TGGCCCTAGC CCCTCAGGTG TGTTCACACC TGTAGCCATT GCCGCTAGGC TTGGCACGGG	60
GAATCCGCG	69

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 45 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TTTTGAATTC CCATTCAGGA CCCAGACCTG AAACCCAGGG AATCC	45
---	----

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TGCCTTGAAG TGCTTCTTCA

20

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCTCAGAGAT GACGAGAATG C

21

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4042 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GTCAACCTTC ACAGTAATTG CTTGTTCACT GACTGCCACA ACCCAGCCTG GCAGAGAGAG	60
GGAAGATACC CTATAAAGCA AGGTAACGTT AATGTTGAGA CCATGAATGG CCTTGAGCAG	120
AGCAGAGTAT CATTGCTTCC TTCAAATTC AGAAGGATCT GATGGTGCTC TGTGAGTTCA	180
TGGGGGTGCC TCCGTGCAGG TTGAAACCAC AGCTGTCGTC CTTCCGCTTT CCCTCTTGAT	240
CAGTAGAAGG GTACCCTCCC TGGCCTGCAC GTCGCTGGGT CACACAACAC TGGCTGTCGT	300
TGCACAAAGC CACGGCCACC AGCGTTCCTT TGAGGCCATT TGTTCCAGC CATGGTGCCT	360
ATAGGATTTT TCCTTTATCC TGTAATTCA GCCAATCAG AGCATGTGAC CTGGCTTAGA	420

TGTCAATATA	ATTGTTGTTA	TGTGCTCTTT	TCCCTTCCTG	TGTCTGTGAC	AGGTTTAATT	480
TAACCTGAGA	AGGCTGCAGA	TCCTCGGGGG	TTGGTGTA	AACACCTCAT	CCTGATCTGA	540
GAAGGCGGTC	AGCTTTTCTC	CTCGTTGCCG	TTGGCTGCCA	GCACCCATTC	TCTGTGGATG	600
TGAAAATCCC	AGAAGGGCTG	GGCTTCCTTC	TTGGCATTCC	CCAGGCCTAT	CTCCAGAGTG	660
GGGCCCAGCA	TGGGAGGATT	GTACCCCACT	CACTCCCCTG	ATGTGGGGCT	TGGACCTACA	720
GCTCGACAGC	ACCCATGGAA	TGTGGGCAGA	AGCGACAGCA	GCCAACGTCC	GCCTTGGCCT	780
TAGGGCGGCA	CGTGTTCTGC	TTGTGCCCTG	GGAGCCTCCA	CCTTCCACAC	TGTGGGAAGA	840
GGGTGCCCAG	GGAGCTGCAG	TCTCTCCAGC	CCAGCCCCAG	GACGAGGCCC	AGGCAGCAGA	900
GCCACCCCAG	CAGACCTGGC	AGTGTGAGAG	AAATGCATGT	GTATACACTG	AGTTTGCAGG	960
TGGCTGTTAC	ATGGCAGCAT	TGACTGACAC	AGACAGAAAA	GAGATCCACG	AGGGAGAAGT	1020
GAGAGTGCTG	GAGACTCCAA	CAAGCCACAG	GCTGCAGGGG	CAGGATGGCT	TCTTAGAAGG	1080
TGAATGATTG	TTCTGGGAAT	CTATCAGAGG	AAGACATAGA	GGCTCCAGAC	GGTTGAAGGC	1140
CCAACAGTGA	TCCCAGACGG	GCCCCATGTC	AGACCAGGCT	CCTCCAGGGC	TGTCGCTGCC	1200
CTCACCAAAG	CCCGTCCTGA	GGGCAGCCAC	ACAGCAGGCA	GCACTCGCCA	TTTGTACAAG	1260
CGAGGCCCAA	GTTCCAGCCT	TCCTTCTGGC	AGGTAGAGGA	AGCAGGGGCA	CTATGCCTGG	1320
GAGTTCCTGA	AAGCAGATGG	GGCAGCATTT	GGTCAAGAGC	CAGGAGGGGA	TGACAGACCA	1380
GAGGGGAACC	CTCGTCCCAC	GTGCTGAGCA	CACGTAGGGG	GTTGGGCACT	TGCTCTGTGA	1440
GCTATAATTG	GTGTCCCTGT	GCCCCGCCGG	AAGCTGCACC	AGGCAGTTTC	TTGGTGGAGG	1500
ACAGTGGCCG	CCCTCTAGCT	TTACTCCCTT	CCCCGTGATG	GGTCGCTGTC	AGATGTGTGT	1560
CCAGGAAAGG	CAAACACCAA	AGGCAGAGGA	CTAGTCCCTA	CACCGAATAC	TCCGGTGGCC	1620
TTGCTTGGGG	GCTGGGTTTT	GACGTGCTGG	AGGCTGTCTT	AGACTTAGAG	ATTAAAAACA	1680
GGGAAGAACC	ATTGCTGAAA	CCTTTGGAAA	AGCCTGCAAT	GGCCTCTGGC	AGCCTGAGGA	1740
GTGGTGGTGT	TTCCATCTGG	TAGACGCCGT	CTCAATAGGA	GGGACAGATG	AGTGCACCAG	1800
TGCTGCCAGC	CAGAGGCGTC	TGTTGGCGTG	TCTTTATGGA	ATGGGGTGCC	AGTCTTGTGG	1860
AGGGTGGTTT	ACCTTCCTGT	TTCTAGTCCC	CACTGGGCCT	GCCTTCTGCT	TCATGCCAGC	1920
TGGCCAGACC	GAGCACTTTC	CTGACTTTCG	ACCTTGCCCC	CTGCTGACTC	TTGCCGTTGA	1980
GGCCTCCTGC	AGACCCCAT	TGTATTCAAT	TCCTGCAGTT	CTCATACCTG	AATCCCGCCT	2040
GGACTTCTGC	CAACCGTTCC	AGGCCCTCCT	CCAGGGGGGA	CCACAGATGC	TACGTGCAGG	2100
GCTGTCCTTG	GAGGGCCAGC	ACAGCCCCTT	CCAAGTGGGC	AAGACCCAGG	GGTGGCTCAA	2160
AAGATAGCTG	TGCCCTAGCC	CTGGAACCTC	TGAATGTTGA	TTTTTGTAGC	AAAAAAGGAC	2220
TTGCAGATGT	GAGTAAAGGC	TGTTGAGATA	AGGACATCCT	CCCTGCTCTC	TGGGAGGACC	2280

CCAAATGCAG	GTGCACAGAT	CTTAAGAAGA	AGAGGCAGAG	ACTGGGGTGA	TGCAGCCACA	2340
ACTAAGGAAA	GCCAAGGATT	GCTGGCAGCC	TGCAGAACT	GGAGGGCAAG	GAGCATCCCC	2400
CAACCGCCCC	GAGCCTCCAG	GAGGCGCAAG	GTCCTACTGA	CTCCCTGACT	TCAGACGTCC	2460
AGTCTCCGGA	ATTTTGAGAG	GATCCATTTC	TGTTATTTTA	AGCAACCAAA	CTTGTGGTAG	2520
TTTCACCACT	CTCAGGAAAT	GAATACGAAT	GGAAAGTCAA	AGATTCCAAG	AAATGAGTGG	2580
CGGGGTGCGG	TGGCTCACAC	TTGTAATCCC	AGCATTGTGCG	GGAAGATTGC	TTGGGCTCAG	2640
GACTTGAGAG	CCTTGTGTCT	GTGAGAACT	TAAAAAATAG	GCTGGGTGCG	ATCGTCACGC	2700
CTGTAATCCC	AGCACTTTGG	GAGGCCGAGG	CAGGCGGATC	ACAAGGTCAC	GAGTTTGAGA	2760
CCAGTGTGAC	CAACATGGTG	AAACCCTGTC	TCTACTAAAA	ATACAAAAAT	TAGCCGGGTG	2820
TGGTGGTGC	TGCCTGTAAT	CCCAGCTACT	CGGGAGGCTG	AGGCAGAAGA	ATTGCTTGAA	2880
CCCAGGAAGC	AGAGGTTGCA	GTGAGCCGAG	ATAGTATTAC	TGCACTCCAG	GCTGGGCAGC	2940
AGAGCAAGAT	TCCGCCTCAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	CTGAGCATGG	3000
TAGCATGCAC	CTGTGGTCCT	CGTACGCCGG	AGGATTGCCT	GAAGCCAGGA	GTTCAAGACC	3060
AGTCTGGACA	AAAGAGCAAG	ACCCCATCTC	TACCAAAAAA	ATTTAAAAAT	TAGCCAGGCA	3120
TGGTGCCGTA	CCCATAGTCT	TAGCTACTCA	GGAGGCTGAG	GAGGGAGGAT	TATCTGAGCC	3180
TGGCGGTTGA	GGCTATAATG	AGCCATGATT	TGGCCACTGC	ACTCCAGCCT	TGGCAACACA	3240
GTGTGAGACC	CTGTCTCAAA	AACAATAAAA	ACCCAAAACA	AAAGAACCAA	GAAATTACTG	3300
GACCTGAGCC	TGGCCTTTAG	CTGCTGCCCT	GCCCTKTGAC	TGGTCACTCG	GATCCCTGGG	3360
CCTAAACACA	CAGCCTATTG	TCTACCTCAA	GAAGGCTCCC	CACTGCTTGG	CTGGCAATTG	3420
GGGTGGCTTT	GCAGGCCCCA	CCTGTCCTGG	CCCCACGGCG	CTGGTGCTGC	AGGCCCCCAC	3480
CACTGCTTGT	TCCGAGCTCC	CCAGCCTCCT	GCAGAGTTGC	CTGCACCTGA	TGGCGATGAA	3540
TCAGGAAGGC	AGGCGTGTCC	TGGGCCACAG	AGCAGTCATG	CTGTCAGCCA	CCAGGGGGCT	3600
CCATTGCAA	CTTTGGATGT	GGCTTTGGCC	TCTTTGTCCA	AAGTGACCTT	GGGGCCCCCA	3660
GACAAGAGAC	AGGGAGACTG	GAGCCCAGCC	CCACCCTCCC	GCACATACCT	GGCCCATCCC	3720
TGCCCTATCC	TGGAAGATGG	GGGCCACCAC	ACGTRCAAGG	GACACGGGAT	AGGAACCTTT	3780
GGCCTTGTTA	TCAGACATTT	TAAACTAAG	TGCAAACGTG	ATTATCAGGT	GCAGTTTTTA	3840
CAGCAGCAAG	AAACCTGTGC	TTACAGAAAG	AAACACGTGC	TAGCAACCCA	CCTATGCGGA	3900
AAGCCACACA	GAGCCATTGT	TTCTGCACT	CTCAGGTGAC	GGCTCACATT	TGCCCCAGGG	3960
AAGGTCACAG	CTGCCTGAAC	TTTAAAACT	CCCAGACACG	CACTGCCTGT	GCAGGATCCG	4020
GAGCCCAGCA	GCACTGCCAG	GG				4042

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 810 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 471..810

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CCTTGAAGTG CTTCTTCAGA GACCTTTCTT CATAGACTAC TTTTTTTTCT TTAAGCAGCA	60
AAAGGAGAAA ATTGTCATCA AAGGATATTC CAGATTCTTG ACAGCATTCT CGTCATCTCT	120
GAGGACATCA CCATCATCTC AGGTGAGCAC CAGGTGGAGT GCCTCTGGGT GACTGGCCGG	180
TTTGGAGCAG GGAGGGAGGC TTAGAGTCTC ATCCTCCAGC AGCGAGTGAG GCGGAGGCTC	240
CAGCGTCCTC CCGGGCGGGT TTTCTGGTGG ATGGAGGAGT GACTCGGGGT CCTCTACGTG	300
GTGCCAGCTG TTTGGCTTTC TGGACGTTGT AGGAAAGGGT TTCCCCGCC TGCGTCCCCC	360
TGACCTTGAG CTCCACCAGC CCCTGCCAGC TGGGCTCCAG AAGGCTGGAG TGCTGTGGCA	420
GGGATGACGT CTCACTTCTG TTATGTCTCT GTGCCCTGTG CTCTCCCAGG ATG AGG	476
	Met Arg
	1
GGC ATG AAG CTG CTG GGG GCG CTG CTG GCA CTG GCG GCC CTA CTG CAG	524
Gly Met Lys Leu Leu Gly Ala Leu Leu Ala Leu Ala Ala Leu Leu Gln	
	5 10 15
GGG GCC GTG TCC CTG AAG ATC GCA GCC TTC AAC ATC CAG ACA TTT GGG	572
Gly Ala Val Ser Leu Lys Ile Ala Ala Phe Asn Ile Gln Thr Phe Gly	
	20 25 30
GAG ACC AAG ATG TCC AAT GCC ACC CTC GTC AGC TAC ATT GTG CAG ATC	620
Glu Thr Lys Met Ser Asn Ala Thr Leu Val Ser Tyr Ile Val Gln Ile	
	35 40 45 50
CTG AGC CGC TAT GAC ATC GCC CTG GTC CAG GAG GTC AGA GAC AGC CAC	668
Leu Ser Arg Tyr Asp Ile Ala Leu Val Gln Glu Val Arg Asp Ser His	
	55 60 65
CTG ACT GCC GTG GGG AAG CTG CTG GAC AAC CTC AAT CAG GAT GCA CCA	716
Leu Thr Ala Val Gly Lys Leu Leu Asp Asn Leu Asn Gln Asp Ala Pro	
	70 75 80
GAC ACC TAT CAC TAC GTG GTC AGT GAG CCA CTG GGA CGG AAC AGC TAT	764
Asp Thr Tyr His Tyr Val Val Ser Glu Pro Leu Gly Arg Asn Ser Tyr	
	85 90 95

AAG	GAG	CGC	TAC	CTG	TTC	GTG	TAC	AGG	CCT	GAC	CAG	GTG	TCT	GCG	G	810
Lys	Glu	Arg	Tyr	Leu	Phe	Val	Tyr	Arg	Pro	Asp	Gln	Val	Ser	Ala		
100						105					110					

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GACATTGATT ATTGACTAGT T 21

(2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TTTAAGCTTC TGCAGAAAAG ACCCATGGAA AG 32

(2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

TGCTCTGGCA CAACAGGTAG 20

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CATAGATGGT CAATGCGGC

19

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 8355 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AGCTTCTGCT TTAGGAAAGT AGAAAAATAA GAGCAAATTA AATCCAAGGT AAGTAAAAAA	60
AAAAAAAAAA AAAAAAAGAA ATAAAAATTA GAGCAGAAAT CAATAAAATT GAAGACAGTA	120
AATCAATAAA GAAAATCAAC ATAAAAAGTC TGGTTCTTGA AAAGATATAT AAAATTGATA	180
AGCATCTACC TAGGATAATT AAGGAAAAAA GACAGAGGAC ACAGATTACT AATATCAAAC	240
ATAAAAGCGG GAACATCACT GCAAATTTTA TAGGCATTGA AAGCGTAATA AAAGAATACT	300
ATAAACTATT CTATAACTAC AAATTTGATA AGTAAATAGA ATGAACCAAT TCCTTGAAAG	360
ACATAATCTG AAAAATGTAA AAAGAAGAAA TAAACAATCT GAATAGCCTA TATCTATTAA	420
ATAAATTGAA TCAGTAATTA ATAACCTCTC AAAACAGGAA GCACAATGCC CAGATGGGTT	480
CACTAGTGAA TTCTATCAAA TATTTAAAGA AAAAAAATT GTATCAACTT TCTACAATCT	540
CTTTCAGAAG ACAGAAGCAG AGGGAATACT TCCTAAATCA TTCAACTAGG CCAGCATTAC	600
CTTAATACCG GAACTAGAAA ATGACATTAC AAGAAAAGAA AACAACAGAC CAATATCTCT	660
CATGAACAAA GATACAAACA TTTTCAACAA AATATTAGCA AAAAGAAATCC AAGAATGTAT	720
CAAAAAATAT ACACCACAAC CAAGTAGAAT TTATTCCAGA TATGTAAGGG TGGTTCAACG	780
TTTGAAAATC AATTAACGTA ATTTGTCCCA TCAACAGGTT AAAGAAGAAA ATCACATGGT	840
CATATTGATA GACACAGAAA AAGCATTTGA CAAAATTTAA CACCCATTCA TGATGCAATC	900

TCTCAGTAAA	CTAGGAATAG	AGGAAAACTT	CCTCAGCTTG	AATGTACCTT	CCTCTCAATT	960
TTGCTATGAA	CCTGAAACTC	CTCTTAAAAA	ATAAAGTTTT	TCATTTAAAA	AGAAAACAAA	1020
AAACATGGAG	GAGCGTTGAT	GTATCTCATT	TTAGACCAAT	CAGCTATGGA	TAGTTAGGCG	1080
ACAGCACAGA	TAGCTGCTGT	ACTTCTGTTT	CTGGCAATGT	TCCAGACTAC	ATTTAAAAAA	1140
TTTTTAATTA	TAGACTTGTA	CTTAATGTTC	AAGAAAAATA	TGAAAATGCT	TTGCCGTGTT	1200
AATGCTACTC	TTTTTTAAAA	AAAACATAAG	TTCAAACCTT	ATTTATATTT	CATTAGTTTT	1260
TTAGCTACTG	TTCTTTTTCT	GTTCTGGGAT	CTCATTCAGA	ATGCCACATT	ACATATAATT	1320
CTCATGTCTC	CTTGGGTTCC	TCTTAGTTTT	GACAGTTCCT	CAGACTTTTC	TTATTTTTGA	1380
TGACCTTGAC	AGTTTTGAGG	AGTACTGGTT	AGATATAGGG	TAATGGTTTT	TAAAGTATAT	1440
TTGTCATGAT	TTATACTGGG	TAAGGGTTTG	GGAGGAAGCC	ATGGGTAAAGT	ACTGTTCTCA	1500
TCACATCATA	TCAAGTTATA	TACCATCAAT	ATTGCCACAG	ATGTTACTTA	GCCTTTTAAT	1560
ATTTCTCTAA	TTTAGTGTAT	ATGCAATGAT	AGTTCCTCTG	TTTCTGAGAT	TGAGTTTCTC	1620
ATGTGTAATG	ATTATTTAGA	GTTTCTCTTT	CATCTGTTCA	AATTTTGTCT	AGTTTTATTT	1680
TTTACTGATT	TGTAAGACTT	CTTTTTATAA	TCTGCATATT	ACAATTCTCT	TTACTGGGGG	1740
TGTTGCAAAT	ATTTTCTGTC	ATTCTATGGC	CTGACTTTTC	TTAATGGTTT	TTTAATTTTA	1800
AAAATAAGTC	TTAATATTCA	TGCAATCTAA	TTAACAATCT	TTCTTTGTG	GTTAGGACTT	1860
TGAGTCATAA	GAAATTTTTC	TCTACACTGA	AGTCATGATG	GCATGCTTCT	ATATTATTTT	1920
CTAAAAGATT	TAAAGTTTTC	CCTTCTCCAT	TTAGACTTAT	AATTCACCTG	AATTTTTTTG	1980
TGTGTATGGT	ATGACATATG	GGTTCCTTTT	TATTTTTTAC	ATATAAATAT	ATTTCCCTGT	2040
TTTTCTAAAA	AAGAAAAAGA	TCATCATTTT	CCCATTGTAA	AATGCCATAT	TTTTTTCATA	2100
GGTCACTTAC	ATATATCAAT	GGGTCTGTTT	CTGAGCTCTA	CTCTATTTAT	CAGCCTCACT	2160
GTCTATCCCC	ACACATCTCA	TGCTTTGCTC	TAAATCTTGA	TATTTAGTGG	AACATTCTTT	2220
CCCATTTTGT	TCTACAAGAA	TATTTTTGTT	ATTGTCTTTT	GGGCTTCTAT	ATACATTTTA	2280
GAATGAGGTT	GGCAAGTTAA	CAAACAGCTT	TTTTGGGGTG	AACATATTGA	CTACAAATTT	2340
ATGTGAAAAG	AAAGTATACC	TTCAACAATAT	TAAGTCTTTT	AGTTCATGAA	TATAGTATGT	2400
CTCTCCGTTT	CTGCATTAAC	TTAGACATTC	ATTAATTTCT	CTCACAATTT	ATAAGTTTAT	2460
TTAGATCTTC	ATTCATTTAA	ATCTTCACTA	ACCTCTCATT	TACAATTTGT	AAGTTTTCTG	2520
GGTAACAGTC	TTGCACTTCT	TTGCCTAGAT	TTATTTCCAA	GTAGATTATT	TTCATACATC	2580
GTCTATGGTG	TCATTTTAA	AATGTAATTT	TTCACCTTTT	TATTGCTAAA	GAGAGATGAC	2640
TGATTGTTAA	TATTGATCTT	GTGCGTGGCG	ACCTTGCTGA	ATTCTAATCG	TTTATCTATA	2700
AATTCTTTTG	TATTTTGAAT	GTAAACAATT	AGATCATCTG	CATATAATTT	TTAAATCTG	2760

CTTAAATGA	TGTATTTAA	AGGAAGAAAT	TTTAACCCAT	TCATAGGTGA	GCTTCTGCCA	4680
AGATTACTAC	TAATCCTCAG	GAGAAGGGGT	AGAGGAGAAA	CTCCATAAAG	GCAACTGGAA	4740
GTGGAGTATT	AGGAAGCACC	TCAAGAACAC	AATAGCAGGA	AGTAGCTAGA	GAACAAAGAG	4800
AAGAAAACCA	GAAAAAATA	ATCCCTTTTT	ATTTTCTGT	TTCCATTCCCT	TTGGCTCCAT	4860
TTCCACAGCT	ATGGCCTTTA	TTTTCACCCT	CCACAGCCAT	GAGAGCCTCT	GGGCAGGAGT	4920
TCTCCTCGCC	TCTCCCTGTT	CCAATCACCT	CTAACATTTT	TGCCTATTGT	TCTGCCCAGG	4980
GAAAAAACTC	CAGTCTCTTC	TCTGTCAAAG	ACCTCTTGAA	TTAAGTCCAA	ATGCTACACT	5040
CTGGCATTCA	AGACTCGTAA	TACAGCTCAA	CCTGACTTTT	CCACCCTCAG	CCTCCTTGAT	5100
TCCTAAAATG	AAGCCTGTCC	ACAATTGAAG	CTCCTTGTCT	TTGCTCCTGC	AAATTTGTTC	5160
ATTCTCCTGG	CTGTGTTTGT	GCTGGTCTCT	GTCTATCTAG	AGCTGTGGAT	ATCATGGTAT	5220
CTATTGTCTA	TCATGCTAGC	CATGAACCAC	ATGTGGCTGG	TGAGCATTTT	ATATGGTACT	5280
AGTCTAAATT	GACATCTACT	GTGAGTGTA	AAATGTGCAT	TATGTTTTGA	AGACTGTACA	5340
CAAAAATTTAA	TTATCTCATG	AATAATTTTA	GATTGGTTAT	ATGTTGAAAT	TATAATATTT	5400
TGGATATACT	ATGCTAAATA	AAACATATTA	TTAAAATTAA	CTTCACCTGT	TTCTTTTCCT	5460
CTTTCATAT	GGCTACTAGA	GCTTTTAAA	TTGCATTATG	TGACTTTATT	GGACAGTACC	5520
GATTGAATGC	CCTCAACCAC	ATCACCTCAC	CACAGCCACC	TCTACCTGTA	GTGATCATAC	5580
CACTTCTTTA	GGCACACTGC	CTGCATTAAG	GGCAATGAAT	GCCTTTTCAT	CTTCTCCACT	5640
AGATGTAGTT	TCTTTTTTCT	TTGAGAGCCA	TCATCACCAT	CATGGTTGAC	ACCATGAACC	5700
TATCTGAAGA	TGTCAGCCAT	AGACTGCTTG	ATATTCTACA	GGAAAGATCA	CAGTTTTAAG	5760
TGCAATCTAC	CCATGTTATT	AGCAGTGTGT	ATCTTTCACA	CATTACACAG	CCTCTCTAAG	5820
CCTCATTTCT	CTCCTCTGTA	AGATGGGGAT	GATAATAACC	CATCTCAAAT	GTTTACTATG	5880
AGGATTATTC	AAAGAATGGC	AAATAGCAAG	TGCTTAATAA	ATGATAACTA	GTACTACCGC	5940
CACTACTGTT	GTTTTTATTG	TATTAGATTA	TGAACTCTCT	AAGGACCATT	TCCGGATGGA	6000
GGATAAGAGA	CCATTTGATG	TGGGCAGTGA	TGAGGCCTTC	TGTTGCACCT	GGAAAGGTCA	6060
ACTATATACA	AGCCTGCAAG	TCATTCTATA	GGAGCAGGCC	CCAGTGACCA	GACTCTATAG	6120
ACTGTCTCCT	CTTTCCTGAG	AGGGACAGCC	ATCTCTAGGT	TGACTAACCT	CTGAAGCTCC	6180
TTGCATTGGC	TTTTGTGCTA	TGAGCCATGG	ATGATTCCAG	ACTAATCCGA	GAATGCTCGT	6240
CAAAACCCCA	AGGAATTACT	CAAATACTGA	CATAACAGAC	ATTTTTGAGT	GGAAGAGCCG	6300
AGTTTTTTTT	AATATTCTGA	AACTCATTGT	TTTTAAAATG	CATGAGATGG	CCAAGGTCTT	6360
GCTAAGAGCT	GGCCTGCAAA	GCGAAAGGCA	GAGAGAATGA	AACCCATAGA	GAGGCAGAAT	6420
AACCAGAAAG	GTTGGGACTC	GTTTATTTTA	TAATGTAAAT	TAGTCTATTA	TGAAACAATA	6480

CTTGTTTACT	GGTGGAAAAAT	TGGAAAAATAC	AAAGAATAAA	AGGAGGAAAA	AAATCACTCT	6540
TTAGTTTCAC	AAGCCAAAATC	AAGCCACTAT	TAAAATGGTG	GTTTACTTCC	TTTTATTAAT	6600
TTTCTGTACA	TATTTTTGCA	TAATCATGTT	GTATGTACAA	TTTTATGTTT	TATTTTTCAA	6660
TATTAAGTGG	TGTCTTTCAA	ATTTCCCTAAT	GACAAAAATA	ATATATGCTC	ATAATAGAAC	6720
ATTTTAAATG	CAAAATAAAAC	AAAATAAATG	TTAAAATTTA	GTAATATTTA	TTAAATTTTC	6780
TCCAAGTGCA	CGAAATTACA	AATGTAACAA	CCTAATTCCC	TAGTGGCCTA	ATAACCCTAT	6840
TTCCAGACCT	CTTCTCATT	CAAGGAAAAA	CTCATATGCA	GATAGTTCTA	AAGGTATGAA	6900
GTGAAAAGAT	AAAGATTTTT	CTTCCTTGCT	GCATCCTCAC	CCCATCAGCA	TTATTCCCCA	6960
GGGTAAGTAC	TATTAATAGA	TAGTAATTCT	ACCCAAAGGA	AAAAATCATA	TGCATATAAC	7020
AGCATCATAT	GTATACCTTT	CTAGTAACTT	ACAAAACAAA	TGATAATATC	ATATCCCTTC	7080
TTATGTGTAT	TGCTCTTTTC	ACTAAATGTA	TCTGTGATAT	GTGTCTATAT	CAGCTGATTG	7140
TCCTTTTTGA	TGGCTGAATA	ATATTCCATC	TTGTCCACGT	GATAGTATTA	CTTGACAAGC	7200
TCCCTGCTGA	TGGACATTTG	TCTTTGTTAC	TATGATAGTA	ATATAATCAA	CATTTATATA	7260
TGTTTTGTAT	GTATCTATAA	TACACATGCA	CATACACATG	CATATTTCTG	CAGGGATAGC	7320
CATAGTAAAT	AACTAGTAAC	GGTATTGCAA	GTTAAAGGAA	CAATCTCATT	GCTTGAAATT	7380
TTAAATTTTG	AAATACACTG	CCAATTTTCA	TGGTCTCTCC	TTGTAAGCTA	GTTTGGGCTT	7440
TCTCACAGCA	TGACAGGCTC	AGGGCAGTCA	GACCATCCTG	GCCAAAGAGC	AGAGTGCCAC	7500
AGACCACAAC	TGCTTCTAAT	CAGCCATCTT	CCCAAAGCCT	TCTCTTTTTT	CTATTAATAA	7560
CTTTGTATGA	GATTCCATCT	TAATACTTTT	CTGTTGTTTG	GTCTTGTAAG	AGCTTATTTT	7620
TCTGAACCAG	GAAGTGTTTC	AGGGCGGTTT	TTCTAACTTC	ACAGAGCTCC	CTCTTCTGTT	7680
AGCTTTTGTG	AAATGGTCAA	AAACATAGCA	GCCTGCCTTC	TGAGTTCTCC	ATCCCAACCT	7740
GGTTGGGCCT	TCTCTATCCT	TGTCTGTGTT	GTTTATATCC	TGCTGAAGTG	TGATTCCACT	7800
TGTGCAGTTT	CTCCTCTGTG	TAGGATCAAA	AGGGCTGTGG	CTGGTTGGTT	TGAAAATTTT	7860
TTATACCCTA	GACTATTCCA	GTGCCTTTCA	GAAGTTTCCA	AGGCCCTCTC	ACACTAATCT	7920
ATTATCATAT	TGGGCAAAAC	TCCTGTCAGT	TTCAGTACT	ATTCCCTGAT	TGACTTTTCA	7980
GTAAATCTAT	CTCTCAGTCT	TTCAGTATCC	AAAGAAGATT	GGTTCTAGGA	CCACCATCCC	8040
GCTGCCTCCA	CAGATACCAA	AATCAGAGGA	TGCTCAATTC	CCTCTTATAA	AACGTTGCAG	8100
TATTTGCATA	TAATCTGCAC	ATGTATTTCT	GTATATTTTA	AATCATCCCT	AGATTACTTA	8160
TAATACCTGA	TACAATATAA	ATGCTAAATA	GCTGTAACAC	TGTATCTTTA	AAATTTACAT	8220
TATTTTTTGT	TGTTGTATTA	TTATTTTTAT	TGTATTTTTA	AAAAATATTT	TCCATCTACA	8280
GTCAGTAGAA	TCCACGGATA	CAGAACCTAT	GGATAGGAAG	GACCAACTGT	ATCTTTTAGT	8340

GTTTTGAGGT TCTTG

8355

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1584 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

- (ix) FEATURE:
 (A) NAME/KEY: mat_peptide
 (B) LOCATION: 357..917

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AATTCTCAGG TCGTTTGCTT TCCTTTGCTT TCTCCCAAGT CTTGTTTAC AATTGCTTT	60
AGTCATTCAC TGAACTTTA AAAACATTA GAAACCTCA CAGTTGTAA ATCTTTTCC	120
CTATTATATA TATCATAAGA TAGGAGCTTA AATAAGAGT TTAGAAACT ACTAAAATGT	180
AAATGACATA GGAAACTGA AAGGGAGAAG TGAAAGTGGG AAATTCCTCT GAATAGAGAG	240
AGGACCATCT CATATAAATA GGCCATACCC ACGGAGAAAG GACATTCTAA CTGCAACCTT	300
TCGAAGCCTT TGCTCTGGCA CAACAGGTAG TAGGCGACAC TGTTCTGTGT GTCAAC	356
ATG ACC AAC AAG TGT CTC CTC CAA ATT GCT CTC CTG TTG TGC TTC TCC	404
Met Thr Asn Lys Cys Leu Leu Gln Ile Ala Leu Leu Leu Cys Phe Ser	
1 5 10 15	
ACT ACA GCT CTT TCC ATG AGC TAC AAC TTG CTT GGA TTC CTA CAA AGA	452
Thr Thr Ala Leu Ser Met Ser Tyr Asn Leu Leu Gly Phe Leu Gln Arg	
20 25 30	
AGC AGC AAT TTT CAG TGT CAG AAG CTC CTG TGG CAA TTG AAT GGG AGG	500
Ser Ser Asn Phe Gln Cys Gln Lys Leu Leu Trp Gln Leu Asn Gly Arg	
35 40 45	
CTT GAA TAC TGC CTC AAG GAC AGG ATG AAC TTT GAC ATC CCT GAG GAG	548
Leu Glu Tyr Cys Leu Lys Asp Arg Met Asn Phe Asp Ile Pro Glu Glu	
50 55 60	
ATT AAG CAG CTG CAG CAG TTC CAG AAG GAG GAC GCC GCA TTG ACC ATC	596
Ile Lys Gln Leu Gln Gln Phe Gln Lys Glu Asp Ala Ala Leu Thr Ile	
65 70 75 80	
TAT GAG ATG CTC CAG AAC ATC TTT GCT ATT TTC AGA CAA GAT TCA TCT	644
Tyr Glu Met Leu Gln Asn Ile Phe Ala Ile Phe Arg Gln Asp Ser Ser	
85 90 95	
AGC ACT GGC TGG AAT GAG ACT ATT GTT GAG AAC CTC CTG GCT AAT GTC	692
Ser Thr Gly Trp Asn Glu Thr Ile Val Glu Asn Leu Leu Ala Asn Val	
100 105 110	

TAT CAT CAG ATA AAC CAT CTG AAG ACA GTC CTG GAA GAA AAA CTG GAG Tyr His Gln Ile Asn His Leu Lys Thr Val Leu Glu Glu Lys Leu Glu 115 120 125	740
AAA GAA GAT TTC ACC AGG GGA AAA CTC ATG AGC AGT CTG CAC CTG AAA Lys Glu Asp Phe Thr Arg Gly Lys Leu Met Ser Ser Leu His Leu Lys 130 135 140	788
AGA TAT TAT GGG AGG ATT CTG CAT TAC CTG AAG GCC AAG GAG TAC AGT Arg Tyr Tyr Gly Arg Ile Leu His Tyr Leu Lys Ala Lys Glu Tyr Ser 145 150 155 160	836
CAC TGT GCC TGG ACC ATA GTC AGA GTG GAA ATC CTA AGG AAC TTT TAC His Cys Ala Trp Thr Ile Val Arg Val Glu Ile Leu Arg Asn Phe Tyr 165 170 175	884
TTC ATT AAC AGA CTT ACA GGT TAC CTC CGA AAC TGAAGATCTC CTAGCCTGTG Phe Ile Asn Arg Leu Thr Gly Tyr Leu Arg Asn 180 185	937
CCTCTGGGAC TGGACAATTG CTTCAAGCAT TCTTCAACCA GCAGATGCTG TTTAAGTGAC	997
TGATGGCTAA TGTACTGCAT ATGAAAGGAC ACTAGAAGAT TTTGAAATTT TTATTAAATT	1057
ATGAGTTATT TTTATTTATT TAAATTTTAT TTTGGAAAAT AAATTATTTT TGGTGCAAAA	1117
GTCAACATGG CAGTTTAAAT TTCGATTTGA TTTATATAAC CATCCATATT ATAAAATTGC	1177
CAAGTACCTA TTAGTTGTTC TTTTAAAT ATACCTGCAA AGTAGTATAC TTTCTGGCCC	1237
CTGCCTTTAA GGAATTTAAA ATTCAAGAAA GCCATGATGG AATATATAAG GTAAGAGACA	1297
ATAAGGGGAC CTGAACCTTA TGGGGGAATA AATATGGCAT GAACCTGCTGT GGGATTAAAA	1357
GAGAAAAGGA AAGCTGGAGG GTCTGGAACCT AAACCTGGGG TTCCCATTCC TCCTACTGTG	1417
TGTTCCAGAT TCTCTCATCA TAAAGTTAGA ATTGAGCTGG CCATCAGGAA TAGCCAGAGG	1477
AATATGTCAG CTTTGTGTT CTCCCTAACC TTCCCCAGTT ATTTGGGGGA TCACTTTGCT	1537
CCTCGAAAGA TTTTAAATA ATTATGTGCC CCCCACCATC CCTGCAA	1584

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TGACATAGGA AACTGAAAG G

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TTTGGATCCG TTGACAACAC GAACAGTGTC G

31

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TTTCCCGGGA CATTGATTAT TGACTAGTT

29

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 21 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CGTGTCAAGG ACGGTGACTG C

21

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 113 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

Met Arg Gly Met Lys Leu Leu Gly Ala Leu Leu Ala Leu Ala Leu
 1 5 10 15
 Leu Gln Gly Ala Val Ser Leu Lys Ile Ala Ala Phe Asn Ile Gln Thr
 20 25 30
 Phe Gly Glu Thr Lys Met Ser Asn Ala Thr Leu Val Ser Tyr Ile Val
 35 40 45
 Gln Ile Leu Ser Arg Tyr Asp Ile Ala Leu Val Gln Glu Val Arg Asp
 50 55 60
 Ser His Leu Thr Ala Val Gly Lys Leu Leu Asp Asn Leu Asn Gln Asp
 65 70 75 80
 Ala Pro Asp Thr Tyr His Tyr Val Val Ser Glu Pro Leu Gly Arg Asn
 85 90 95
 Ser Tyr Lys Glu Arg Tyr Leu Phe Val Tyr Arg Pro Asp Gln Val Ser
 100 105 110
 Ala

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 187 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Met Thr Asn Lys Cys Leu Leu Gln Ile Ala Leu Leu Leu Cys Phe Ser
 1 5 10 15
 Thr Thr Ala Leu Ser Met Ser Tyr Asn Leu Leu Gly Phe Leu Gln Arg
 20 25 30
 Ser Ser Asn Phe Gln Cys Gln Lys Leu Leu Trp Gln Leu Asn Gly Arg
 35 40 45
 Leu Glu Tyr Cys Leu Lys Asp Arg Met Asn Phe Asp Ile Pro Glu Glu
 50 55 60
 Ile Lys Gln Leu Gln Gln Phe Gln Lys Glu Asp Ala Ala Leu Thr Ile
 65 70 75 80
 Tyr Glu Met Leu Gln Asn Ile Phe Ala Ile Phe Arg Gln Asp Ser Ser
 85 90 95
 Ser Thr Gly Trp Asn Glu Thr Ile Val Glu Asn Leu Leu Ala Asn Val
 100 105 110
 Tyr His Gln Ile Asn His Leu Lys Thr Val Leu Glu Glu Lys Leu Glu
 115 120 125

Lys Glu Asp Phe Thr Arg Gly Lys Leu Met Ser Ser Leu His Leu Lys
130 135 140

Arg Tyr Tyr Gly Arg Ile Leu His Tyr Leu Lys Ala Lys Glu Tyr Ser
145 150 155 160

His Cys Ala Trp Thr Ile Val Arg Val Glu Ile Leu Arg Asn Phe Tyr
165 170 175

Phe Ile Asn Arg Leu Thr Gly Tyr Leu Arg Asn
180 185